HU-HHU at GermEval-2017 Sub-task B: Lexicon-Based Deep Learning for Contextual Sentiment Analysis

Behzad Naderalvojoud
DFG SFB 991
Hacettepe University
Ankara, Turkey
n.behzad@hacettepe.edu.tr

Behrang Qasemizadeh
DFG SFB 991
Universität Düsseldorf
Düsseldorf, Germany
zadeh@phil.hhu.de

Laura Kallmeyer
DFG SFB 991
Universität Düsseldorf
Düsseldorf, Germany
kallmeyer@phil.hhu.de

Abstract

This paper describes the HU-HHU system that participated in Sub-task B of GermEval 2017, Document-level Polarity. The system uses 3 kinds of German sentiment lexicons that are built using translations of English lexicons, and it employs them in a neural-network based context-dependent sentiment classification method. This method uses a deep recurrent neural network to learn context-dependent sentiment weights to change the lexicon polarity of terms depending on the context of their usage. The performance of the system is evaluated using the benchmarks provided by the task's organizers.

1 Introduction

Sentiment lexicons are known as useful language resources in sentiment analysis systems that determine the sentimental orientation of terms out of context. As manually creating such lexicons is expensive and time-consuming, one possible solution is to translate English resources into other languages (Waltinger, 2010; Ucan et al., 2016). However, natural language is ambiguous and word by word translation cannot achieve satisfying results; terms can possess more than one sentiment and can express different sentiments with respect to the context. In this case, sense-based sentiment lexicons, like SentiWordNet (SWN) (Esuli and Sebastiani, 2006), assign polarities to word-senses instead of words. However, the number of word-senses cannot necessarily be the same for two terms that are translations of each other. Moreover, not all word-senses of a term may express subjectivity in a language (Akkaya et al., 2009). Therefore, we generated a sentiment lexicon for German that takes into account SWN synsets. To map subjective synsets to German terms, we extended the approach used in the HUMIR[1] project (Naderalvojoud et al., 2017) for applying it to German language. This approach creates a cross-lingual sense mapping between the SWN synsets and German terms and produces a single polarity value for each term. This value indicates the strength of subjectivity according to the number of mapped synsets. In fact, the polarity and the number of English synsets associated with each term constitute the domain of German terms' sentiments.

Besides this lexicon, we also employ two other German sentiment lexicons proposed in (Waltinger, 2010) that are translations of English Subjectivity Clues (Wilson et al., 2005) and SentiSpin (Takamura et al., 2005) lexicons. Three online English-to-German translation softwares have been used in constructing the German lexicons; a German term that appears in most of translation results is selected as a translation of the given English term. While the polarity values of German Subjectivity clues are assigned manually, they are inherited from the corresponding English resource for German SentiSpin.

The sentiment lexicons generated is employed in a context-dependent sentiment analysis system that uses a deep recurrent neural network (RNN) to capture the contextual sentiment modifications that can change the prior known sentiments of terms with respect to the context. After describing the construction of the proposed German SentiWordNet lexicon in Section 2, we explain the sentiment analysis system we used in Section 3. Section 4 reports the evaluation results. Finally, Section 5 presents the conclusion.

2 German SentiWordNet Lexicon

The proposed German SentiWordNet lexicon is constructed by the following three main steps using the Open Subtitle Corpus (Tiedemann, 2009) for

[1] http://humir.cs.hacettepe.edu.tr/projects/tsa.html

the German–English language pair: (1) We first generate a cross-lingual distributional/statistical model to represent subjective English terms[2] in the German language vector space. In this model, each English term is represented by a distributional semantic vector whose elements are German terms (Naderalvojoud et al., 2017). (2) The generated model is used to represent the SWN synsets. To represent synsets, the semantic vectors of the synset terms are summed up. (3) Synset mapping is applied to the reduced semantic vectors[3] for mapping German terms to subjective SWN synsets. We suppose that German terms with high frequency in the semantic vector are more likely associated with the given synset.

As a result of this approach, we achieved two lexicons of 14,309 (named SWN1) and 43,790 (named SWN2) German subjective terms by using two different corpora having 70,534 and 13,883,398 movie subtitles, respectively.

3 Context-Dependent Sentiment Analysis Using Deep Learning

We use deep learning to capture the implicit sentiment knowledge contained in the semantic/syntactic structure of a sentence. In fact, the prior sentiment of terms in the lexicon can be changed based on the negation, intensification, or semantic structure of terms in the context. For example, the positive sentiment of “good” is shifted to negative in the sentence “Nobody gives a good performance in the team” by the word “nobody”. A similar situation can be seen for the word “great” in the sentence “He was a great liar” and its positive sentiment (based on the lexicon) is shifted to negative. Thus, the use of sentiment lexicons without consideration of the context cannot achieve a satisfying result.

In the sentiment analysis method we use, the contextual sentiment knowledge is combined with the terms’ prior sentiments. To this end, we employ a context-sensitive lexicon-based method proposed in (Teng et al., 2016). In this approach, the sentiment score of a sentence is computed based on the weighted sum of the polarity values of the subjective terms obtained from the lexicon. The learned weights are considered as context-dependent features that modify the prior polarity values of terms with respect to the context. Therefore, the effects of negation, shifting and intensification can be considered in the sentiment classification task. The overall structure of the model is simply shown in Eq. 1.

$$Score = \sum_{k=1}^{N} \gamma_k \times LexScore(t_k) + b \qquad (1)$$

In Eq. 1, N denotes the number of subjective terms in the sentence, $LexScore(t_k)$ is the polarity value of term t_k in the lexicon, γ_k is the context-dependent weight of term t_k and b is the sentence bias score.

To Learn γ and b, following (Teng et al., 2016), we employ a recurrent neural network (RNN) model with bidirectional long-short-term-memory (BiLSTM) cells (Graves et al., 2013; Sak et al., 2014) for extracting the semantic composition features.

We use the sentiment annotated data provided by the task organizers (Wojatzki et al., 2017) for training the model. The German FastText pre-trained word embeddings[4] are used in generating the model in combination with the German sentiment lexicons. We tune hyper-parameters of our model using the obtained classification results on the development set.

4 Evaluation Report

We evaluate the proposed sentiment model using customer reviews about “Deutsche Bahn” provided by the task organizers[5]. Table 1 shows the statistics of data in three categories, “positive”, “negative” and “neutral”. In order to show the significance of contextual sentiment analysis, we also indicate the contextual ambiguity of the training set by relying on the occurrence of subjective terms in irrelevant reviews. For example, in Table 1, while column “*NegInPos*” shows the percentage of positive reviews with negative clue terms, the column “*PosIn-Neg*” indicates the occurrence of positive clue terms in negative reviews. The most important point is the occurrence of subjective terms in the neutral reviews (column “*SubInNeu*”) which can make it hard to distinguish neutral reviews from subjective ones. The other observation is that the number of neutral reviews outnumbers the positive and nega-

[2]A term is subjective if it has at least a synset with non-zero positive or negative polarity value. SWN includes 29,095 subjective synsets; the number of synset terms belonging to these subjective synsets is 39,746.

[3]For synset mapping, we reduced the dimension of vectors to 10 by selecting the most frequent terms.

[4]https://github.com/Kyubyong/wordvectors

[5]https://sites.google.com/view/germeval2017-absa/data

tive ones. For example, only 6% of train reviews are positive, whereas 68% are neutral.

Table 1: Statistics of dataset

Data Split	All	Pos 6%	Neg 26%	Neu 68%
train	19432	1179	5045	13208
dev	2369	148	589	1632
test-syn	2566	105	780	1681
test-dia	1842	108	497	1237
sum	26209	1540	6911	17758
		Contextual ambiguity on the train set %		
German Lexicon		NegInPos	PosInNeg	SubInNeu
SWN1		100.00	93.18	100.00
SentiSpin		93.64	96.13	98.37
SubjectivityClue		98.05	72.80	98.73

As the train set is imbalanced, the performance of the classification model can tend towards the majority class (Naderalvojoud et al., 2015). As a result of this, the micro F-measure value (which is the shared task evaluation metric) is affected by the majority class. Hence, we use the macro F-measure value along with the micro score in the evaluation.

Table 2 indicates the best results on the development set achieved from the shared task baseline system (SVM) and our context-dependent system (RNN). We select the model that produces the best results on the development set for applying to test set. As two lexicons generated from SWN achieve the best results in comparison to two other lexicons, we constructed two models according to these lexicons for testing.

From the results shown in Table 2, the RNN model outperforms the baseline system in all three classes. While the baseline system yields weak F-measure values for positive and negative classes, the RNN-based system achieves F-measure values of 0.4533 and 0.6254. Despite the lack of positive instances in the train set (6%), the RNN model can achieve much better results than SVM in combination with the proposed German SWN lexicons. This can be also observed in the negative class in which the F-measure value increases from 0.2212 in SVM to 0.6254 in RNN. Overall, the change in Positive–Negative macro F-measure value from 0.1173 (in Baseline-SVM) to 0.5394 (in SWN1-RNN) clearly shows the effect of the proposed lexicon-based context-dependent sentiment analysis method. It is worth noting that a German lexicon has been also used in the baseline system, however, this system has not made an impact on the sentiment classification result as much as the proposed RNN model has made. Furthermore, while the German SentiSpin lexicon does not improve the performance of the RNN model in the positive class, the proposed German SWN lexicons significantly improve its performance. Although the German subjectivity clue lexicon performs better than SentiSpin, the proposed German SWN1 and SWN2 outperform it by 7% and 4% in MacroF1(PN), respectively.

As the neutral class is the majority class in the train set, all systems yield high F-measure values for this class. Nevertheless, the RNN model in combination with two German SWN lexicons achieves the best results in terms of macro F-measure value (0.6452, SWN1-RNN) and micro F-measure value (0.7873, SWN2-RNN) over all classes.

Tables 3 and 4 show the results of the baseline and proposed systems on the synchronic and diachronic test sets, respectively. From these results, we can observe that the proposed system outperforms the baseline method by using all German sentiment lexicons. In synchronic test set, while the SWN1-RNN achieves the best macro F-measure value (0.4907), SWN2-RNN yields the best micro F-measure value (0.7494). In diachronic test set, the RNN model with both German SWN lexicons achieves the best micro F-measure values. However, they do not maintain this superiority and the RNN model with German Subjectivity clue lexicon gives the best macro F-measure value (0.5211). This may arise from the fact that the polarity values of German Subjectivity Clues are manually assigned. As a result, the proposed context-dependent sentiment analysis system performs well in combination with the German SWN lexicons and remarkably outperforms the baseline SVM model.

5 Conclusion

This paper presented the sentiment analysis approach of HU-HHU system in the GermEval 2017 shared task. In this approach, an RNN model is used to learn the context-dependent sentiment weights that can change the lexicon polarity of terms depending on the context. As shown in the empirical evaluations, compared to the baseline system, this approach significantly improves the performance of the sentiment classification task.

References

Cem Akkaya, Janyce Wiebe, and Rada Mihalcea. 2009. Subjectivity word sense disambiguation. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-*

Table 2: The best results on the development set

Lexicon-Model	F1pos	F1neg	F1neu	MacroF1(PN)	MacroF1(all)	MicroF1
Baseline-SVM	0.0134	0.2212	0.8244	0.1173	0.3530	0.7092
SWN1-RNN	**0.4533**	**0.6254**	0.8568	**0.5394**	**0.6452**	0.7847
SWN2-RNN	0.4381	0.6086	**0.8602**	0.5234	0.6356	**0.7873**
SentiSpin-RNN	0.0127	0.6075	0.8525	0.3101	0.4909	0.7716
SubjectivityClues-RNN	0.4112	0.5932	0.8591	0.5022	0.6212	0.7838

Table 3: Results on synchronic test set

Lexicon-Model	MacroF1	MicroF1
Baseline-SVM	0.3325	0.6730
SWN1-RNN	**0.4907**	0.7366
SWN2-RNN	0.4806	**0.7494**
SentiSpin-RNN	0.4764	0.7159
SubjectivityClues-RNN	0.4718	0.7357

Table 4: Results on diachronic test set

Lexicon-Model	MacroF1	MicroF1
Baseline-SVM	0.3539	0.6894
SWN1-RNN	0.5165	**0.7362**
SWN2-RNN	0.5036	**0.7362**
SentiSpin-RNN	0.4482	0.7176
SubjectivityClues-RNN	**0.5211**	0.7323

Volume 1, pages 190–199. Association for Computational Linguistics.

Andrea Esuli and Fabrizio Sebastiani. 2006. SENTIWORDNET: A high-coverage lexical resource for opinion mining. *Institute of Information Science and Technologies (ISTI) of the Italian National Research Council (CNR).*

Alex Graves, Abdel-rahman Mohamed, and Geoffrey Hinton. 2013. Speech recognition with deep recurrent neural networks. In *Acoustics, speech and signal processing (icassp), 2013 ieee international conference on*, pages 6645–6649. IEEE.

Behzad Naderalvojoud, Ebru Akcapinar Sezer, and Alaettin Ucan. 2015. Imbalanced text categorization based on positive and negative term weighting approach. In *International Conference on Text, Speech, and Dialogue*, pages 325–333. Springer.

Behzad Naderalvojoud, Alaettin Ucan, and Ebru Akcapinar Sezer. 2017. A novel approach to rule based turkish sentiment analysis using sentiment lexicon. *The Scientific and Technological Research Council of Turkey (TÜBİTAK), 115E440.*

Haşim Sak, Andrew Senior, and Françoise Beaufays. 2014. Long short-term memory recurrent neural network architectures for large scale acoustic modeling. In *Fifteenth Annual Conference of the International Speech Communication Association.*

Hiroya Takamura, Takashi Inui, and Manabu Okumura. 2005. Extracting semantic orientations of words using spin model. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pages 133–140. Association for Computational Linguistics.

Zhiyang Teng, Duy-Tin Vo, and Yue Zhang. 2016. Context-sensitive lexicon features for neural sentiment analysis. In *EMNLP*, pages 1629–1638.

Jörg Tiedemann. 2009. News from opus-a collection of multilingual parallel corpora with tools and interfaces. In *Recent advances in natural language processing*, volume 5, pages 237–248.

Alaettin Ucan, Behzad Naderalvojoud, Ebru Akcapinar Sezer, and Hayri Sever. 2016. SentiWordNet for new language: Automatic translation approach. In *Signal-Image Technology & Internet-Based Systems (SITIS), 2016 12th International Conference on*, pages 308–315. IEEE.

Ulli Waltinger. 2010. GERMANPOLARITYCLUES: A lexical resource for German sentiment analysis. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC)*, Valletta, Malta, May. electronic proceedings.

Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the conference on human language technology and empirical methods in natural language processing*, pages 347–354. Association for Computational Linguistics.

Michael Wojatzki, Eugen Ruppert, Sarah Holschneider, Torsten Zesch, and Chris Biemann. 2017. Germeval 2017: Shared task on aspect-based sentiment in social media customer feedback. In *Proceedings of the GSCL GermEval Shared Task on Aspect-based Sentiment in Social Media Customer Feedback.*